



PATENT ABSTRACTS OF JAPAN

(11) Publication number: **07014000 A**

(43) Date of publication of application: 17.01.95

(51) Int. Cl.

G06K 9/36
G06K 9/62

(21) Application number: 05154763

(22) Date of filing: 25.06.93

(71) Applicant: **MATSUSHITA ELECTRIC IND CO LTD**(72) Inventor: **KAMIMURA YUJIRO**(54) **TABLE RECOGNIZING DEVICE**

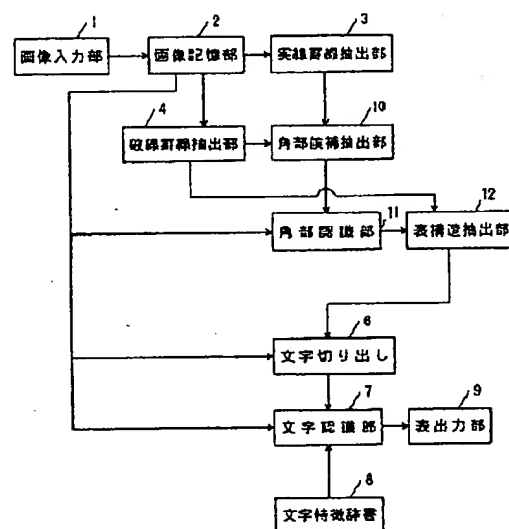
(57) Abstract:

PURPOSE: To provide a table recognizing device with excellent reliability which is capable of rightly recognizing a table structure even if an angular part is composed of circular arc, oblique direction ruled lines, etc.

CONSTITUTION: A table recognizing device is provided with a run extraction part extracting runs in the longitudinal and lateral direction from binary image or data, a solid-line ruled line extraction part 3 extracting runs which are a solid line threshold value or more in length of the runs, connecting these and extracting solid-line ruled lines and a broken-line ruled line extraction part 4 extracting the runs of specific patterns which continue in the length to be broken line threshold value or more of the runs, connecting these and extracting broken-line ruled line. This device has an angular part candidate extraction part 10 extracting the part where the endpoint of the longitudinal ruled line and the endpoint of the lateral ruled line are close to each other in the solid-line ruled lines extracted in the solid-line ruled line extraction part 3 and the ruled lines extracted in the broken-line ruled line extraction part 4 as angular part candidates and an angular part recognition part 11

performing the recognition of the the angular parts by matching the part of the angular part candidates extracted in the angular part candidate extraction part 10 with a preliminarily prepared angular part pattern.

COPYRIGHT: (C)1995,JPO



(51)Int.Cl.⁸G 0 6 K 9/36
9/62

識別記号

庁内整理番号

F I

技術表示箇所

Z 8623-5L

審査請求 未請求 請求項の数1 O L (全 7 頁)

(21)出願番号 特願平5-154763

(22)出願日 平成5年(1993)6月25日

(71)出願人 000005821

松下電器産業株式会社

大阪府門真市大字門真1006番地

(72)発明者 上村 裕二郎

大阪府門真市大字門真1006番地 松下電器
産業株式会社内

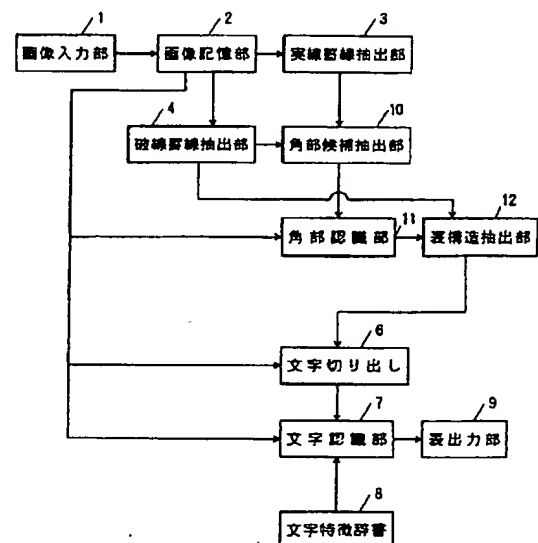
(74)代理人 弁理士 小堀治 明 (外2名)

(54)【発明の名称】 表認識装置

(57)【要約】

【目的】 角部が円弧状・斜方向野線等からなる場合でも表構造を正しく認識することができる信頼性に優れた表認識装置の提供。

【構成】 2値画像かデータから縦・横方向のランを抽出するラン抽出部と、ランの内長さが実線しきい値以上のランを抽出しこれを連結して実線野線を抽出する実線野線抽出部3と、ランの内特定パターンのランが破線しきい値以上の長さ続くものを抽出しこれを連結して破線野線を抽出する破線野線抽出部4と、を備えた表認識装置であって、実線野線抽出部3で抽出された実線野線及び破線野線抽出部4で抽出された野線の内縦方向野線の端点と横方向野線の端点とが近い部分を角部候補として抽出する角部候補抽出部10と、角部候補抽出部10で抽出された角部候補の部分に予め用意された角部パターンとマッチングさせて角部の認識を行う角部認識部11と、を備えた構成を有している。



【特許請求の範囲】

【請求項1】認識対象文書から得られる2値画像データ中の縦・横方向のランを抽出するラン抽出部と、前記ラン抽出部で抽出されたランの内長さが実線しきい値以上のランを実線罫線要素として抽出しこの実線罫線要素を連結して実線罫線を抽出する実線罫線抽出部と、前記ラン抽出部で抽出されたランの内特定パターンのランが破線しきい値以上の長さより続くものを破線罫線要素として抽出しこの破線罫線要素を連結して破線罫線を抽出する破線罫線抽出部と、を備えた表認識装置であって、前記実線罫線抽出部で抽出された実線罫線及び前記破線罫線抽出部で抽出された破線罫線からなる罫線の内縦方向罫線の端点と横方向罫線の端点とが近い部分を角部候補として抽出する角部候補抽出部と、前記角部候補抽出部で抽出された角部候補の部分を予め用意された角部パターンとマッチングさせて角部の認識を行う角部認識部と、を備えたことを特徴とする表認識装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は印刷・手書きされた表を読み取り、その表の表構造と、表の各セル内に記入された文字を認識する表認識装置に関するものである。

【0002】

【従来の技術】近年、コンピュータの普及に伴い、省力化のために、コンピュータ等の入力装置として、キーボードの代わりに、伝票等の表をその表構造を含めて認識することができる表認識装置が用いられるようになった。

【0003】以下に従来の表認識装置について説明する。図6は従来の表認識装置の機能ブロック図である。1はスキャナ等からなり認識対象文書を画像として入力し2値画像データを出力する画像入力部、2は画像入力部1より出力される2値画像データを記憶する画像記憶部、3は画像記憶部2中の2値画像データから縦・横方向の黒画素の並び（以下ランと呼ぶ）を抽出しその長さが実線しきい値以上のランを実線罫線要素として抽出してこの実線罫線要素を連結することで実線罫線を抽出する実線罫線抽出部、4は画像記憶部2中の2値画像データから縦・横方向のランを抽出し特定パターンのランが破線しきい値以上の長さより続くものを破線罫線要素として抽出してこの破線罫線要素を連結することで破線罫線を抽出する破線罫線抽出部、5は実線罫線抽出部3で抽出された実線罫線と破線罫線抽出部4で抽出された破線罫線とからなる罫線の相互関係から表構造を認識して表内の文字等が記入されたセルを抽出する表構造抽出部、6は表構造抽出部5で抽出されたセル内の文字を切り出す文字切り出し部、7は文字切り出し部6で切り出されたセル内の文字と全ての文字の文字特徴を記憶している文字特徴辞書8とを比較してその類似度が文字認識しきい値より高い場合にその文字を認識文字としてそれ

に対応するJISコード等の文字コードを出力する文字認識部、9は表構造抽出部5から出力されたセル等の表構造と文字認識部7より出力される文字コードとを統合して得られる表認識結果を出力する表出力部である。

【0004】以上のように構成された従来の表認識装置について、以下その動作を説明する。図7(a)は認識対象文書の一例を示す図であり、図7(b)は罫線抽出結果の一例を示す図である。aは縦方向の実線罫線と横方向の実線罫線とで区切られた定形セル、bは円弧状罫線からなる角部を有する非定形セルである。

【0005】初めに、図7(a)に示すような認識対象文書を画像入力部1から入力し、得られた2値画像データを画像記憶部2に記憶する。次に、実線罫線抽出部3によって縦方向及び横方向の実線罫線を抽出する。次に、破線罫線抽出部4によって縦方向及び横方向の破線罫線を抽出する。この時抽出される実線・破線罫線を図7(b)に示す。

【0006】ここで、図7(b)より明らかなように、非定形セルbの角部が円弧状罫線よりなるために、縦・横方向罫線のみを抽出する実線罫線抽出部3、破線罫線抽出部4によっては、この部分が抽出されない。そのため、非定形セルbをセルとして認識することができず、表構造を誤認識することがあった。

【0007】

【発明が解決しようとする課題】しかしながら上記従来の構成では、実線罫線抽出部及び破線罫線抽出部が縦・横方向罫線のみしか抽出することができず、角部に円弧状・斜方向罫線等を含む表では、これらの罫線を抽出することができず汎用性に欠けるという問題点を有していた。また、角部にこれらの罫線を有する非定形セルをセルとして認識することができず、表構造を誤認識してしまい非定形セル中に記入された文字等を認識することができず信頼性に欠けるという問題点を有していた。

【0008】本発明は上記従来の問題点を解決するもので、角部が円弧状・斜方向罫線等からなる場合であってもこれらの罫線を認識することができ汎用性に優れ、角部にこれらの罫線を有する非定形セルであっても認識することができ、表構造を正しく認識することができるのと同時に、この非定形セル内の文字等を正確に認識することができる信頼性に優れた表認識装置を提供することを目的とする。

【0009】

【課題を解決するための手段】この目的を達成するために本発明の表認識装置は、認識対象文書から得られる2値画像データから縦・横方向のランを抽出するラン抽出部と、前記ラン抽出部で抽出されたランの内長さが実線しきい値以上のランを実線罫線要素として抽出しこの実線罫線要素を連結して実線罫線を抽出する実線罫線抽出部と、前記ラン抽出部で抽出されたランの内特定パターンのランが破線しきい値以上の長さより続くものを破線

罫線要素として抽出しこの破線罫線要素を連結して破線罫線を抽出する破線罫線抽出部と、を備えた表認識装置であって、前記実線罫線抽出部で抽出された実線罫線及び前記破線罫線抽出部で抽出された破線罫線からなる罫線の内縦方向罫線の端点と横方向罫線の端点とが近い部分を角部候補として抽出する角部候補抽出部と、前記角部候補抽出部で抽出された角部候補の部分と予め用意された角部パターンとマッチングさせて角部の認識を行う角部認識部と、を備えた構成を有している。

【0010】ここで、角部パターンの形状としては、略円弧状、略楕円弧状、略多角形状等が好適に用いられる。

【0011】

【作用】この構成によって、角部候補抽出部が抽出されている実線罫線と破線罫線のうち縦方向罫線の端点と横方向罫線の端点とが近い部分を角部候補として抽出し、角部認識部が予め用意された角部パターンとこの角部候補とをマッチングさせて角部を認識するために、表の角部が円弧状・斜方向罫線等からなる場合であっても、これらの罫線を認識することができる。また、角部にこれらの罫線を有する非定形セルでも認識することができ、表構造を正しく認識することができるとともに、この非定形セル内の文字を正確に認識することができる。

【0012】

【実施例】以下本発明の一実施例における表認識装置について、図面を参照しながら説明する。図1は本発明の一実施例における表認識装置の機能ブロック図であり、図2は本発明の一実施例における表認識装置の装置ブロック図である。1は画像入力部、2は画像記憶部、3は実線罫線抽出部、4は破線罫線抽出部、6は文字切り出し部、7は文字認識部、8は文字特徴辞書、9は表出力部であり、これは従来例と同様なものなので同一の符号を付し説明を省略する。10は実線罫線抽出部3で抽出された実線罫線及び破線罫線抽出部4で抽出された破線罫線からなる罫線の内縦方向罫線の端点と横方向罫線の端点とが近い部分を角部候補として抽出する角部候補抽出部、11は角部候補抽出部10で抽出された角部候補の部分と予め用意された略円弧状、略多角形状等の角部パターンとマッチングさせて角部の認識を行う角部認識部、12は実線罫線抽出部3で抽出された実線罫線及び破線罫線抽出部4で抽出された破線罫線並びに角部認識部11で認識された角部から認識対象文書の表構造を抽出するとともに定形セル及び非定形セルを抽出する表構造抽出部である。図2において、13は表認識装置全体を制御するCPU、14は画像記憶部2を構成する画像データ領域15と2値画像データ中のランを記憶するラン領域16と表認識の結果が得られる表構造、抽出された定形・非定形セル内に記入された文字に対応するJISコード等のコード情報からなる表データを記憶する表データ領域17とを有するRAM、18は文字特徴辞書

8が記憶された文字特徴辞書領域19と角部認識部11で用いられる角部パターンが記憶された角部パターン領域20とCPU13への指示が記載されたプログラムが記憶されているプログラム領域21とを有するROM、22は画像入力部1を構成するスキャナ、23は利用者が表認識装置へ認識開始の指示等を与えるためのキーボード、24は表データ領域17中の表データに基づいて認識された表等を表示するための表示装置、25は上記のデバイス間を接続するバスである。

【0013】以上のように構成された本発明の一実施例における表認識装置について、以下その動作を説明する。図3は本発明の一実施例における表認識装置の全体フローチャートであり、図4は本発明の一実施例における表認識装置の角部抽出処理の詳細なフローチャートであり、図5(a)は認識対象文書の一例を示す図であり、図5(b)は罫線抽出結果の一例を示す図であり、図5(c)は角部認識処理結果を示す図であり、図5(d)は認識対象文書の他の例を示す図である。aは定形セル、bは非定形セルであり、これらは従来例と同様なものなので同一の符号を付し説明を省略する。図5(d)において、cは斜め方向罫線からなる角部を有する非定形セルである。

【0014】初めに、画像入力部1によって、図5

(a)に示すような認識対象文書を読み込み、画像記憶部2に2値画像データとして記憶する(S1)。次に、画像記憶部2中に記憶されている2値画像データの傾きを補正する(S2)。次に、画像記憶部2中に記憶されている2値画像データから、縦方向及び横方向のランを抽出する(S3)。次に、実線罫線抽出部3によって、S3で抽出された縦方向及び横方向のランの内、長さが実線しきい値以上のランを実線罫線要素として抽出する(S4)。次に、実線罫線抽出部3によって、S4で抽出された実線罫線要素を連結して実線罫線を抽出する(S5)。次に、破線罫線抽出部4によって、S3で抽出された縦方向及び横方向のランの内、特定パターンのランが破線しきい値以上続くものを破線罫線要素として抽出する(S6)。次に、破線罫線抽出部4によって、S6で抽出された破線罫線要素を連結して破線罫線を抽出する(S7)。ここで、上記の処理によって、図5

(b)に示すような罫線が抽出される。次に、角部候補抽出部10及び角部認識部11によって、円弧状罫線等からなる角部を抽出する(S8)。このstepの詳細は後述する。ここで、この処理によって、図5(c)に示すように角部を含む全ての罫線を抽出することができる。次に、表構造抽出部12によって、S5、S7で抽出された実線・破線の罫線及びS8で抽出された角部を基にして、表構造を抽出し、個々のセルを抽出する(S9)。この処理によって、図5(c)に示すように、縦方向、横方向罫線のみからなる定形セルaだけでなく、円弧状罫線からなる角部を有する非定形セルbも正しく

抽出することができる。次に、セル内の文字認識が全て終了したか調べる（S10）。Noである場合は、文字切り出し部6、文字認識部7によって、個々の定形セルa、非定形セルb内に記入された文字を認識してそれに対応するJISコード等の文字コードを出力して、次のセルを認識するためにS10へjumpし（S11）、Yesである場合は、表出力部9によって、S9で抽出された表構造及びS11で認識された文字コード等の表認識結果に基づいて認識された表を表示装置24に表示する（S12）。ここで、上記の処理によって、図5

(a)に示すような表を正確に認識することができる。
【0015】次に、S8における角部抽出処理について説明する。まず、図4において、角部候補抽出部10によって、S5、S7で抽出された罫線から、端点に近い縦方向罫線と横方向罫線とのペアを抽出する（S13）。次に、S13で抽出されたペアの内、処理が済んでないものがあるか調べる（S14）。Noである場合は、角部抽出処理を終了し、Yesである場合は、角部認識部11によって、S13で抽出された縦方向罫線と横方向罫線のペアの交差点の2値画像データを切り出し、予め用意された角部パターンとマッチングする（S15）。次に、S15でのマッチングの結果得られる類似度が角部しきい値よりも大きいか調べる（S16）。Noである場合はS14へjumpし、Yesである場合は、角部認識部11によってその部分を角部として出力する（S17）。次に、この角部を有する非定形セルb内に記入された文字の認識を容易にするため、角部の部分の画像を消去し、次の角部を認識するために、S14へとjumpする（S18）。

【0016】尚、本実施例においては、図5(a)に示すような角部に円弧状罫線を有する非定形セルbを含む表を認識する場合について説明したが、角部パターンを変更することによって、図5(d)に示すような角部に斜方向罫線を有する非定形セルcを含む表等であっても、同様に正しく認識することができる。

【0017】

【発明の効果】以上のように本発明は、角部候補抽出部が、抽出されている実線罫線と破線罫線のうち縦方向罫線の端点と横方向罫線の端点とが近い部分を角部候補として抽出し、角部認識部が、予め用意された角部パターンとこの角部候補とをマッチングさせて角部を認識するために、表の角部が円弧状・斜方向罫線等からなる場合であっても、これらの罫線を認識することができ汎用性に優れ、角部にこれらの罫線を有する非定形セルでも認識することができ、表構造を正しく認識することができるとともに、この非定形セル内の文字等を正確に認識することができる信頼性に優れた表認識装置を実現できるものである。

【図面の簡単な説明】

【図1】本発明の一実施例における表認識装置の機能ブロック図

【図2】本発明の一実施例における表認識装置の装置ブロック図

【図3】本発明の一実施例における表認識装置の全体フローチャート

【図4】本発明の一実施例における表認識装置の角部抽出処理の詳細なフローチャート

10 【図5】(a)は本実施例における認識対象文書の一例を示す図

(b)は本実施例における罫線抽出結果の一例を示す図

(c)は本実施例における角部認識処理結果を示す図

(d)は本実施例における認識対象文書の他の例を示す図

【図6】従来の表認識装置の機能ブロック図

【図7】(a)は従来技術における認識対象文書の一例を示す図

(b)は従来技術における罫線抽出結果の一例を示す図

20 【符号の説明】

1 画像入力部

2 画像記憶部

3 実線罫線抽出部

4 破線罫線抽出部

5 表構造抽出部

6 文字切り出し部

7 文字認識部

8 文字特徴辞書

9 表出力部

30 10 角部候補抽出部

11 角部認識部

12 表構造抽出部

13 CPU

14 RAM

15 画像データ領域

16 ラン領域

17 表データ領域

18 ROM

19 文字特徴辞書領域

40 20 角部パターン領域

21 プログラム領域

22 スキャナ

23 キーボード

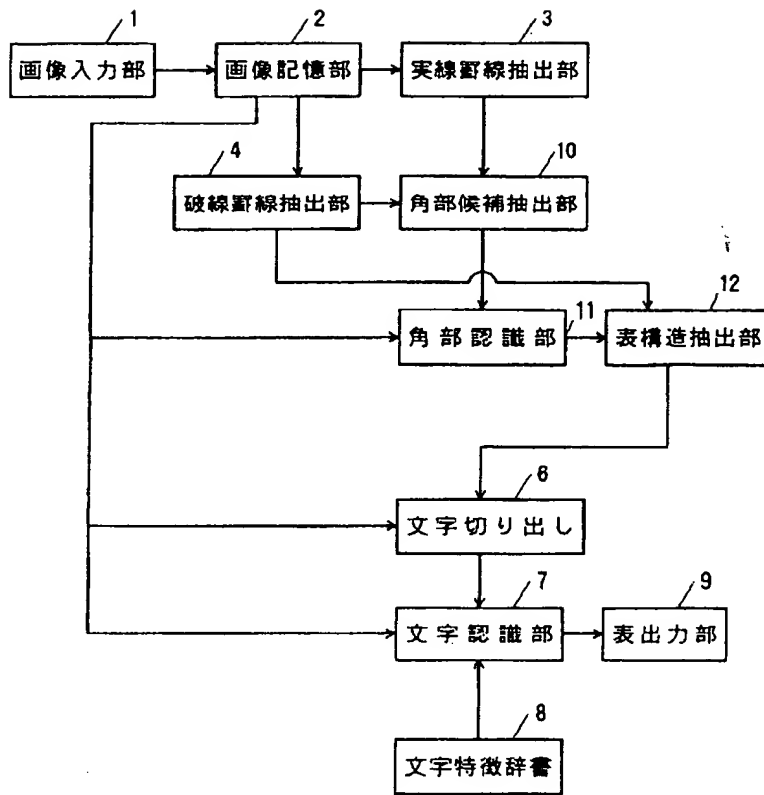
24 表示装置

25 バス

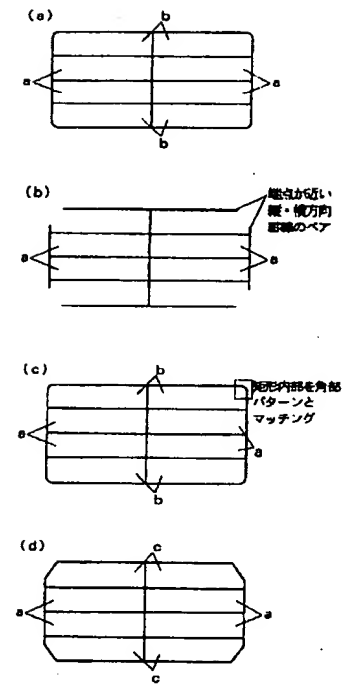
a 定形セル

b, c 非定形セル

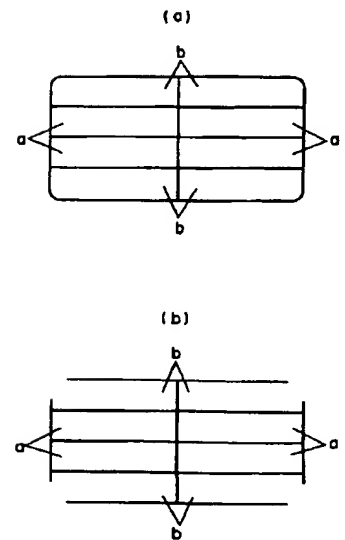
【図1】



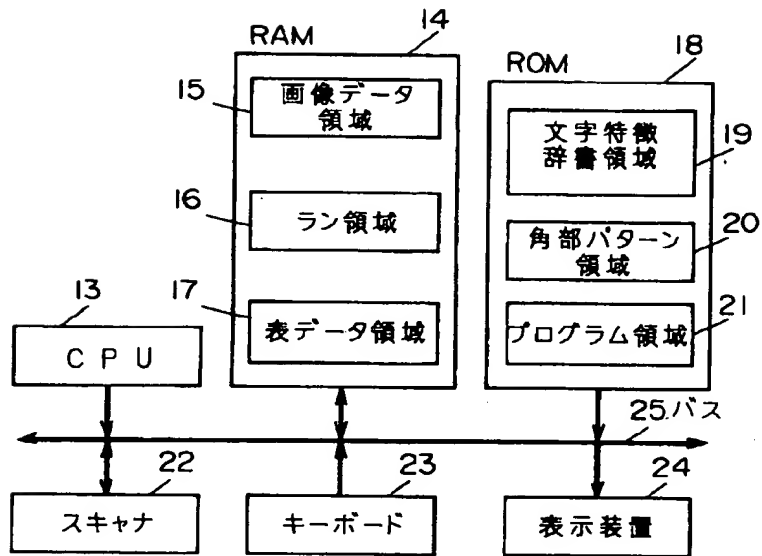
【図5】



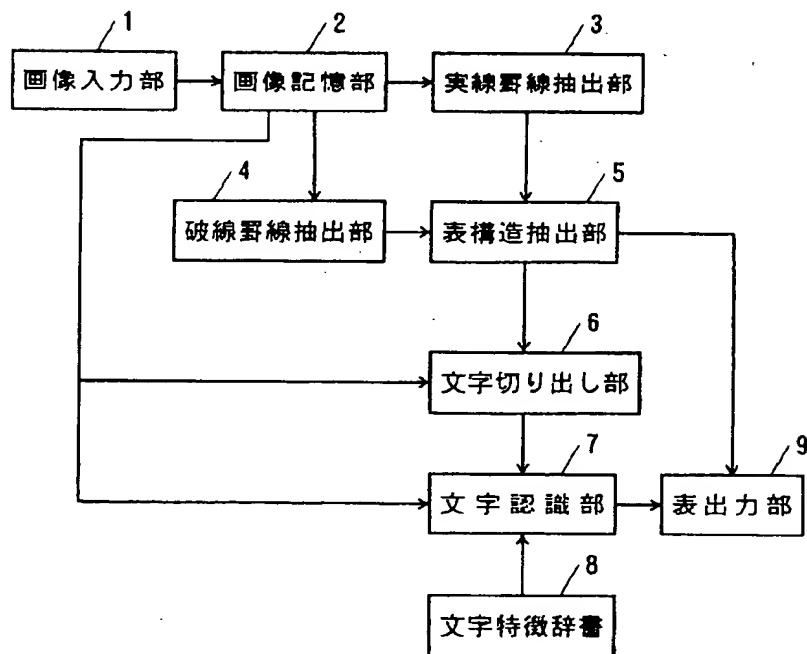
【図7】



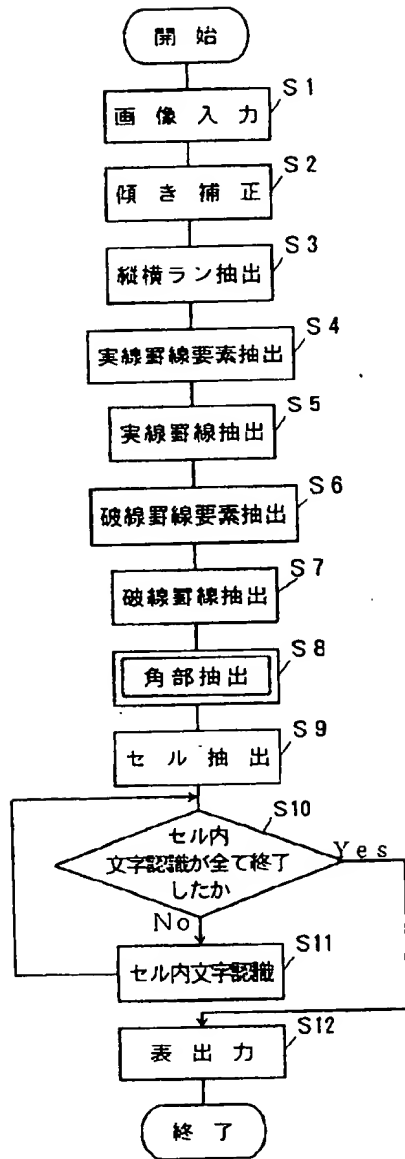
【図2】



【図6】



【図3】



【図4】

